Systematic Gene Search in the Incyte LifeSeq Database

Normal tissue
~50,000 individual ESTs

Tumor tissue
~50,000 individual ESTs

Priority list
High

    Prostate

    Breast

    Ovary

    Bladder

    Uterus

Low

Iterative assembling
with
increasing mismatch

~8,000 contigs
+
~25,000 individual
sequences

~8,000 contigs
+
~25,000 individual
sequences

Comparison of databases

normal tissue-
specific
(expected: 100-500)

nonspecifically
expressed genes

tumor tissue-
specific
(expected: 100-500)

Genes of Interest

Figure 1

REPLACEMENT PAGE (RULE 26)

~50,000 ESTs of a tissue (e.g.: uterus tumor)

GAP4 Assembly 1st Round:
minimum initial match: 20
maximum number of inserted
   empty spaces per sequence: 8
maximum percent mismatch: 0

GAP4 Database 1   unassembled
Contigs 1   ESTs
Individual sequences 1

GAP4 Assembly 2nd Round:
minimum initial match: 20
maximum number of inserted
   empty spaces per sequence: 8
maximum percent mismatch: 1

GAP4 Database 2   unassembled
Contigs 2   ESTs
Individual sequences 2

GAP4 Assembly 3rd Round:
minimum initial match: 20
maximum number of inserted
   empty spaces per sequence: 8
maximum percent mismatch: 2

GAP4 Database 3:   unassembled
Contigs 3   ESTs
Individual sequences 3

Figure 2b1

REPLACEMENT PAGE (RULE 26)

GAP4 Database 3:   unassembled
Contigs 3   ESTs
Individual sequences 3

Consensus 3   GAP4 Assembly 4th Round:
                                     minimum initial match: 20
                                     maximum number of inserted
                                        empty spaces per sequence: 8
                                     maximum percent mismatch: 2

GAP4 Database 4:   unassembled
Contigs 4   ESTs
Individual sequences 4

Consensus 4   GAP4 Assembly 5th Round:
                                     minimum initial match: 20
                                     maximum number of inserted
                                        empty spaces per sequence: 8
                                     maximum percent mismatch: 4

GAP4 Database 5:   unassembled
Contigs 5   ESTs 5
Individual sequences 5

Consensus 5   Individual sequences 5

Figure 2b2

REPLACEMENT PAGE (RULE 26)

Consensus 3   Individual sequences 5

Consensus 4   unassembled
                               ESTs 5

Consensus 5

GAP4 Assembly 6th Round:
minimum initial match: 20
maximum number of inserted
   empty spaces per sequence: 8
maximum percent mismatch: 4

Assembled database
of a specific tissue
(e.g.: uterus tumor)

Figure 2b3

REPLACEMENT PAGE (RULE 26)

Assembled database
of a specific tissue
(e.g.: uterus tumor)

Consensus 6

Read-in as individual sequences

Database
of a specific tissue
(e.g.: uterus tumor)

Database of a second
specific tissue
(e.g.: normal uterus)

GAP4 Assembly
minimum initial match: 20
maximum number of inserted
  empty spaces per sequence: 8
maximum percent mismatch: 4

Tumor tissue-
specific ESTs

Non-tissue-
specific ESTs

Normal tissue-
specific ESTs

Fig. 2b4

REPLACEMENT PAGE (RULE 26)

In silico subtraction of gene expression in various tissues

~30,000 consensus sequences
normal tissue

~30,000 consensus sequences
tumor tissue

Assembly at 4% mismatch

Normal tissue
Specific genes

Cancer tissue
Specific genes

Genes expressed in both tissues

Figure 3

REPLACEMENT PAGE (RULE 26)

Genes of interest

Determination of tissue-specific expression
via electronic Northern (INCYTE LifeSeq and
public EST databases)

Candidate genes for tumor suppressors or
tumor activators

Figure 4a

09/646778

Partial cDNA sequence
e.g., EST or contig
S

...GCCTCAAGTTATC...

WHILE $C_i > C_{i-1}$

Electronic Northern Blot

Fisher's Exact Test   IF H_0
                                    EXIT

Automatic Lengthening

Consensus sequence C

...ATGTCCTAGCCTCAAGTTATCAGATGCAA...

Figure 4b

REPLACEMENT PAGE (RULE 26)

09/646778

Isolation of genomic BAC and PAC clones

Chromosomal clone localization via FISH

Hybridization signal

Sequencing of clones that are located in regions that have chromosomal deletions in prostate and breast cancer leads to identification of candidate genes

Exon   Intron

Confirmation of candidate genes by screening of mutations and/or deletions in cancer tissues

Figure 5

REPLACEMENT PAGE (RULE 26)